Online database for brain cancer-implicated genes: exploring the subtype-specific mechanisms of brain cancer

Min Zhao
University of the Sunshine Coast
Yining Liu
University of the Sunshine Coast
Guiqiong Ding
Beijing Institute of Technology
Dacheng Qu
Beijing Institute of Technology
Hong Qu (quh@mail.cbi.pku.edu.cn)
Peking University

Research Article

Keywords: brain cancer, database, genetic, subtype, systems biology, bioinformatics

Posted Date: March 4th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-269016/v1

Abstract

Background: Brain cancer is one of the eight most common cancers occurring in people aged 40+ and is the fifth-leading cause of cancer-related deaths for males aged 40-59. Accurate subtype identification is crucial for precise therapeutic treatment, which largely depends on understanding the biological pathways and regulatory mechanisms associated with different brain cancer subtypes. Unfortunately, the subtype-implicated genes that have been identified are scattered in thousands of published studies. So, systematic literature curation and cross-validation could provide a solid base for comparative genetic studies about major subtypes.

Results: Here, we constructed just such a literature-based brain cancer gene database (BCGene). In the current release, we have a collection of 1,421 unique human genes gathered through an extensive manual examination of over 6,000 PubMed abstracts. We comprehensively annotated those curated genes to facilitate biological pathway identification, cancer genomic comparison, and differential expression analysis in various anatomical brain regions. When we compared those implicated genes between different subtypes, we found subtype-specific genetic events that had high mutational frequencies. Finally, gene prioritization helps users determine the relative importance of the curated genes, and top-ranked genes were significantly associated with survival rates in a combined dataset of more than 2,000 cancer cases.

Conclusion: BCGene provides a useful tool for exploring the genetic mechanisms of and gene priorities in brain cancer. BCGene is freely available to academic users at http://soft.bioinfo-minzhao.org/bcgene/.

Background

Brain cancer, a leading type of cancer that causes death in both children and adults, was diagnosed in about 300,000 new cases and caused 241,000 deaths globally in 2018 [1]. More recently, mortality figures of brain and other nervous system cancers in the United States caused an estimated 23,890 deaths in 2020 (12,590 males and 10,300 females) [2]. Histologically, glioma is the most common tumor type and includes astrocytoma, ependymoma, and oligodendroglioma. Oligodendroglioma is more sensitive to chemotherapy than is astrocytoma, and therefore has a better overall prognosis [3]. The overall 5-year survival rate of brain cancer patients is approximately 36%, but the 5-year survival rate of oligodendroglioma patients is about 80.6%, and the 10-year relative survival rate is 63.8%. However, the 5-year survival rate for patients with glioblastoma (also known as glioblastoma multiforme, or GBM) is only 5.4%, and the 10-year survival rate is only 2.7% [4]. Therefore, exact identification of glioma subtypes is essential for neuro-oncologists to provide the best treatment. Although many existing clinical and histological methods identify brain cancer subtypes, molecular subtype information can independently and reliably confirm or refute those identifications, thus providing more accurate diagnostic evidence.

Brain cancer is a heterogeneous disease, and the complex molecular mechanisms of its uncontrolled cell growth may be caused by promoter methylation, deregulated gene expression, and/or genetically altered

tumor-suppressor genes and oncogenes [5]. Although thousands of published articles have focus on brain cancer subtypes, a literature-based effort that scrutinizes both the common and unique genetic information of each brain cancer subtype does not exist. Additionally, most functional or clinical studies have been single-gene–based, and thus have failed to provide any comprehensive descriptions of tumorigenesis for different cancer subtypes. Therefore, we developed a database, BCGene, that is a reusable genetic resource for brain cancer, has links to the appropriate literature, and provides global genetic profiles of brain cancer subtypes. The curated genes in the literature can be prioritized according to their correlations with brain cancer, and common and unique cellular events in different brain cancer subtypes can be identified.

Materials And Methods

Literature search and curation

We relied heavily on the PubMed and GeneRIF (Gene Reference Into Function) databases to assemble our collection of brain cancer-implicated genes [6]. Specifically, in the GeneRIF database, we performed a keyword-based query using a Perl regular expression to extract relevant sentences we had previously described [7]: “[gG]liomas or [gG]lioblastomas or [Bb]rain tumor or [Bb]rain cancer or [Aa]strocytomas or [Oo]ligodendrogliomas or [Ee]pendymomas or [Mm]eningiomas or [Hh]aemangioblastomas or [Aa]coustic neuromas or [Cc]raniopharyngiomas or [Ll]ymphomas or [Hh]aemangiopericytomas or [Ss]pinal cord tumor or [Nn]euroectodermal tumor or [Mm]edulloblastoma or [Pp]ituitary tumor”. In total, within 2,881 unique PubMed abstracts, we found 9,304 short sentences related to brain cancer. We used the same expression to search the PubMed database, and all matching records from PubMed and GeneRIF were merged to remove redundancies. Further literature curation included clustering abstracts, extracting matching cancer subtypes, collecting species information, and formalizing gene symbols. For example, in the sentence “re-expression of N-cadherin in gliomas restores cell polarity and strongly reduces cell velocity, suggesting that loss of N-cadherin could contribute to the invasive capacity of tumour astrocytes”, N-cadherin is a common alias for the gene *CDH2* in the Human Gene Nomenclature Database. We also collected tumor subtypes, such as “gliomas”. For non-human genes, we mapped all genes to human orthologous genes. In total, we curated 1,421 human protein-coding genes (Table S1).

Biological annotation and pre-calculated data

To provide biological insight for those collected genes, we retrieved comprehensive biological functional annotations from public resources as described previously [8]. In addition, we used The Cancer Genome Atlas (TCGA) large-scale database to calculate genomic mutation information. For example, the resulting copy number gains and losses in TCGA-GBM and TCGA low-grade glioma (LGG) will enable investigation of changes at the thousands-of-bases level, which may have been overlooked by those published studies focusing on the single nucleotide mutations. We also mapped our 1,421 genes to the gene expression

information from all brain regions in the most updated Allen Human Brain Atlas, thus providing potential gene expression patterns for hundreds of anatomical locations.

The web interface

Based on a systematic survey of genes implicated in brain cancer in the literature, we developed a web interface to make those annotations publicly available. From our web interface, curated subtype information allows users to explore all brain cancer-implicated genes, and the amount of literature evidence for each gene provides a guide to how reliably a gene of interest is associated with brain cancer. We also built a responsive, mobile-friendly webpage by using a Bootstrap framework to provide a grid-based layout.

As shown in Figure 1A, three search modules are implemented by entering 1) a gene name or its description; 2) a gene ontology, (including biological processes), molecular function, and cellular component; and 3) any keywords of interest in the curated literature. These keyword-based queries enables users to identify both curated genes and the related literature on a specific biological topic. For advanced bioinformatics analysis, users may download curated genes, applicable literature, and subtypes in bulk (Figure 1B). To organize information for each gene, we divided our annotation details into six categories: gene information, published evidence, gene ontology, biochemical pathway, genetic mutation summary from TCGA, and gene expression information from the Allen Brain Map (Figure 1C).

Functional enrichment analysis

We used ToppFun [9] to conduct a functional enrichment analysis of the 44 genes shared by multiple subtype groups. In that analysis, we used all 1,421 genes in our BCGene database as background and then used the hypergeometric model, comparing the differences between the 44 annotated genes and all 1,421 genes, to identify the statistical significances of enriched annotations. Since we calculated thousands of raw *p*-values, we then used the Benjamini-Hochberg multiple correction method to adjust those raw values. Focusing on the most significant changes, we extracted the enriched annotations with corrected *p*-values less than 0.01 and used them as over-representative annotations for the 44 genes. Finally, we visualized those enriched biological process terms by the TreeMap package using R language.

Gene prioritization based on functional similarity

Since we have 833 genes with only a single study in the literature, we had to consider the relative importance of each gene when ranking candidate genes according to their functions. To accomplish this, we first built a gold standard, brain cancer gene list that we subsequently used to train an algorithm to identify important functional features. The training gene list included the 27 most reliable genes, each of which was supported by more than 20 published studies in the literature. By using the gene ranking tool

ToppGene [9], we extracted those genes' functional features (e.g., gene ontology, human phenotype ontology, protein domains, biological pathways, known protein-protein interactions, binding transcription factors, co-expression patterns, disease annotations, and data mined from the literature) and used them to build a gene ranking model. We then used that model to calculate each gene's functional relevance.

Cancer genomic analysis of the 33 top-ranked genes that are mentioned in only one published article

We input the 33 genes that have only one published study into cBioPortal to obtain a summary pattern across multiple brain cancer datasets [10]. Then, using the OncoPrint module in cBioPortal, we visualized the sample-based mutational patterns of 2,997 brain cancer samples from 14 studies. To provide the most comprehensive mutational profile, we included the most important genetic mutations in cancer development and progression: single nucleotide variations, gene fusions, and copy number variations (CNVs) [11-13]. We also used mutually exclusive analyses as an overview for mutational complementary patterns across all the samples. Finally, we plotted the correlations between mRNA expression and copy number variant/methylation for each gene of interest and conducted an overall survival analysis of the 2,997 patient samples found with at least one of those 33 genes.

Results And Discussion

The literature frequency for various brain cancer subtypes

Based on our comprehensive literature curation, we cleaned up all the associations between brain cancer genes and the literature before conducting further analyses. As shown in Figure 2A, we found 27 genes that were each supported by more than 20 PubMed abstracts. However, 883 of the 1,421 genes implicated in brain cancer (62%) were supported by only a single evidentiary mention in the literature; so obviously, those genes' functions need further experimental validation. Using cancer subtype keywords, we assigned the 1,421 genes to different subtypes, while a gene could be associated with multiple cancer subtypes, each subtype has its own literature-based evidence (Table S2). As shown in Figure 2B, the top three keywords were: glioma (associated with 582 genes), lymphoma (associated with 450 genes), and medulloblastoma (associated with 245 genes). To explore the genetic heterogeneity of brain cancer, we grouped curated subtype information. For example, astrocytoma, oligodendroglioma, ependymoma, GBM, LGG, ganglioglioma, and oligoastrocytoma were all grouped as gliomas, and medulloblastoma was grouped with neuroectodermal tumors. Then, we subsequently identified 809 glioma-related genes and 354 neuroectodermal tumor-related genes in those two major subtype groups.

After we curated 227 and 25 genes for GBM and LGG, respectively, we summarized all the GBM and LGG CNVs on the gene pages in BCGene. To demonstrate how well our data identifies potential tumor suppressors and oncogenes, we first identified 85 GBM-associated tumor suppressors with more copy number loss (the ratio between copy number loss and copy number gain > 2.0) and 39 GBM-associated

oncogenes with more copy number gain (the ratio between copy number gain and copy number loss > 2.0). Then, by cross mapping to the tumor suppressor and oncogene databases (TSGene 2.0 [14] and ONGene [7], respectively) (Figure 2C), we found that 23 GBM genes with more frequent copy number loss are known tumor suppressor genes, and another 15 GBM genes with more frequent copy number gain are known oncogenes.

Functional enrichment of those genes shared by different subtype groups

To check the genetic heterogeneity of the high-level cancer subtype groups, we overlapped their associated genes to compare the common and unique genetic features of the five subtype groups (glioma, lymphoma, meningioma, neuroectodermal tumor, and pituitary tumor) (Figure 3A) and found 44 genes belonging to four or more groups. Gene ontology enrichment analysis revealed that those 44 genes are highly associated with 12 functional categories (Figure 3B). Some of those categories are highly related to cancer, such as negative regulation of programmed cell death (Benjamini and Hochberg false discovery rate (FDR) corrected *p*-value = 4.35E-05), DNA metabolism regulation (Benjamini and Hochberg FDR corrected *p*-value = 1.42E-04), and regulation of the mitotic G1/S transition (Benjamini and Hochberg FDR corrected *p*-value = 3.79E-04). A most interesting finding was the response to hypoxia (Benjamini and Hochberg FDR corrected *p*-value = 3.31E-04). In general, hypoxia is important in drug resistance and poor survival [15]. Therefore, targeting hypoxia might be a practical way to improve patient survival rate of patients with astrocytoma and GBM [16].

KEGG pathway analysis further highlighted a few important cancer-related signaling pathways, such as the PI3K-Akt signaling pathway (corrected *p*-value = 8.04E-05), pathways in cancer (corrected *p*-value = 5.32E-10), proteoglycans in cancer (corrected *p*-value = 3.33E-06), and the advanced glycation end products-receptor for advanced glycation end products pathway (corrected *p*-value = 1.201E-5). More interestingly, signaling by interleukins (corrected *p*-value = 3.7E-05) and cytokine signaling in the immune system (corrected *p*-value = 1.06E-03) highlighted the importance of interleukins in the progression of brain cancer. Previous observations confirmed that many cytokines (mainly interleukins) are involved in brain cancer aggressiveness and the generation of disease-associated pain [17]. In summary, all our functional analyses demonstrated that subtype-specific gene mining using the BCGene database may be used to identify common genes in different brain cancer subtypes and to explore potential common molecular mechanisms.

Potential prognostic applications

To further explore potential prognostic applications of curated brain cancer-implicated genes, we overlapped the 44 shared genes with 18 brain cancer datasets that have survival outcomes and that are in the prognostic database PRECOG [18] (Figure 3C). Those datasets were grouped into two categories:

12 related to glioma and 6 related to non-glioma. For each gene, PRECOG calculated *z*-scores that characterized gene expression features and clinical outcomes. In general, a positive *z*-score for a gene related to a specific dataset means higher expression and adverse survival, while a negative *z*-score reflects lower expression and favorable survival. By clustering the *z*-scores, all genes could be ordered into three clusters. We then used signal-to-noise ratios (the ratio of the level of a desired signal to the level of background noise) to compare each gene between the glioma and non-glioma groups. In the first group, *PTGS2* had the best signal-to-noise ratio (0.63), meaning that *PTGS2* more powerfully shows signals than noise, making it more useful to distinguish the glioma and non-glioma groups. In contrast, *TP53* in the second cluster had a negative signal-to-noise ratio (-0.79), meaning that its signal was lower than its noise. Additionally, in terms of the fold change of the *z*-scores between the two groups, *PTGS2* is 2696.87 while *TP53* is just 0.03, so *PTGS2* may be a better differential prognostic indicator than *TP53* [19]. In summary, these distinguishing links to different subtypes may provide evidence for the distinct mechanisms related to the survival of patients having different cancer subtypes.

Identify top-ranked genes with evidence mentioned only once in the literature

To further explore the curated genes' relevancies to brain cancer, we ranked all the 1421 genes based on the 27 most reliable brain cancer genes as training set. The reliability of these 27 genes are based on each gene having 20 or more evidentiary mentions in the literature. This ranking result is to generate relatively importance to the remaining 1,394 (1421 minus 27) genes in our database (Table S3). With similar functions to the 27 genes in the training set, the subsequent 100 top-ranked genes are likely important in brain cancer development. And within those top-ranked genes, 33 were linked only by a single support from the literature. Thus, we consider that the roles of those 33 genes in brain cancer development are likely underestimated.

To investigate the potential oncogenic roles of those 33 genes, we used the large-scale cancer genomics datasets in cBioportal [10]. Altogether, we combined 2,997 samples from 14 independent studies, including four datasets related to medulloblastoma, two datasets related to glioma, two GBM studies, two LGG studies, a study of anaplastic oligodendroglioma and anaplastic oligoastrocytoma, a study of a brain tumor patient-derived xenograft, an investigation of pilocytic astrocytoma, and a dataset of pheochromocytoma and paraganglioma. As shown in Figure 4, sample-based mutational patterns revealed 536 samples (18% of the total 2,997 samples) that had at least one genetic mutation related to one of the 33 genes. After closely scrutinizing their subtype information (Figure 5A), we found that the 33 genes were highly mutated in the glioma and GBM datasets but had relatively low mutational rates in the four datasets related to medulloblastoma. Interestingly, those 33 genes had a huge effect on patient survival (Figure 5B). Among the 2,303 patients with survival information, 467 of them had one or more genetic mutations in the 33 genes. The median survival of those 467 patients was 24.59 months, but the remaining 1,836 patients' median survival was 42.20 months, a very significant difference (log rank test, p = 2.30E-8).

Among the 536 samples with genetic mutations in one or more of the 33 genes, the top-ranked gene, *CDK4*, was mutated in 202 samples (8% of the 2,997 samples) and the second-ranked gene, *MAP3K1*, was mutated in 79 samples (2.8%), and 8 of those samples also had a *CDK4* mutation. Since the mutated genes in that mutational pattern are almost mutually exclusive, they may have complementary roles in the progression of brain cancer [20]. As shown in Figure 6A, amplified *CDK4* in five samples coincided with mRNA up-regulation, but four of the five samples had low methylation, which could have caused the increased mRNA expression (Figure 6C). However, *MAP3K1*'s correlation patterns were strikingly different than *CDK4*'s (Figure 6B, D). Altogether, *CDK4* provides a good example of consistent mRNA up-regulation based on both amplification and methylation patterns, and *MAP3K1* may be a good candidate for evaluating some brain cancers' progressions, but those possibilities need further study.

Conclusions

By curating thousands of published articles, we collected 1,421 genes associated with various brain cancer subtypes. From our data collection, 809 gliomas, 450 lymphomas, and 354 neuroectodermal tumor-related genes are supported by evidence in the literature. This comprehensive data collection not only presents the genetic heterogeneity of brain cancer, but also provides comparable genetic resources for exploring the common genetic mechanisms among different brain cancer subtypes. Our additional integrative analysis demonstrates that our database, BCGene, is useful for ranking brain cancer genes' relative importance, which may be used to predict patient survival. Our future plans are to focus on the subtype-unique gene sets, which may both aid the understanding of underlying disease mechanisms and identify novel therapies for specific brain cancer subtypes.

Abbreviations

BCGene: Brain Cancer-implicated Genes and Literature

GBM: Glioblastoma

LGG: Low Grade Gliomas

TCGA: The Cancer Genome Atlas

GeneRIF: Gene Reference Into Function

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

All the data is free to use for academic purpose at: http://soft.bioinfo-minzhao.org/bcgene/

Competing interests

The authors declare no conflict of interest.

Funding

This work was supported by the National Key Research and Development Program of China (No. 2017YFC1201200), the National Natural Science Foundation of China (No. 31671375), and the research start-up fellowship of university of sunshine coast to MZ. The funders did not play any role in research design, data collection, analysis and interpretation, and manuscript writing.

Authors' contributions

M.Z., D.Q. and H.Q. designed the project. M.Z. and G.D. collected data. M.Z., Y.L. and H.Q. performed the analysis. H.Q. supervised the project. All authors wrote the manuscript draft, which M.Z. prepared original draft, and H.Q. finalized the manuscript.

Acknowledgements

Not applicable.

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A: **Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries**. *CA Cancer J Clin* 2018, **68**(6):394-424.
2. Siegel RL, Miller KD, Jemal A: **Cancer statistics, 2020**. *CA Cancer J Clin* 2020, **70**(1):7-30.
3. Sepulveda-Sanchez JM, Munoz Langa J, Arraez MA, Fuster J, Hernandez Lain A, Reynes G, Rodriguez Gonzalez V, Vicente E, Vidal Denis M, Gallego O: **SEOM clinical guideline of diagnosis and management of low-grade glioma (2017)**. *Clin Transl Oncol* 2018, **20**(1):3-15.
4. Gittleman H, Boscia A, Ostrom QT, Truitt G, Fritz Y, Kruchko C, Barnholtz-Sloan JS: **Survivorship in adults with malignant brain and other central nervous system tumor from 2000-2014**. *Neuro Oncol*

2018, **20**(suppl_7):vii6-vii16.

5. Binder H, Willscher E, Loeffler-Wirth H, Hopp L, Jones DTW, Pfister SM, Kreuz M, Gramatzki D, Fortenbacher E, Hentschel B *et al*: **DNA methylation, transcriptome and genetic copy number signatures of diffuse cerebral WHO grade II/III gliomas resolve cancer heterogeneity and development**. *Acta Neuropathol Commun* 2019, **7**(1):59.
6. Coordinators NR: **Database resources of the National Center for Biotechnology Information**. *Nucleic Acids Res* 2016, **44**(D1):D7-19.
7. Liu Y, Sun J, Zhao M: **ONGene: A literature-based database for human oncogenes**. *J Genet Genomics* 2017, **44**(2):119-121.
8. Liu Y, Luo M, Jin Z, Zhao M, Qu H: **dbLGL: an online leukemia gene and literature database for the retrospective comparison of adult and childhood leukemia genetics with literature evidence**. *Database (Oxford)* 2018, **2018**.
9. Chen J, Bardes EE, Aronow BJ, Jegga AG: **ToppGene Suite for gene list enrichment analysis and candidate gene prioritization**. *Nucleic Acids Res* 2009, **37**(Web Server issue):W305-311.
10. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, Jacobsen A, Byrne CJ, Heuer ML, Larsson E *et al*: **The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data**. *Cancer Discov* 2012, **2**(5):401-404.
11. Wee Y, Wang T, Liu Y, Li X, Zhao M: **A pan-cancer study of copy number gain and up-regulation in human oncogenes**. *Life Sci* 2018, **211**:206-214.
12. Li X, Liu Y, Lu J, Zhao M: **Integrative analysis to identify oncogenic gene expression changes associated with copy number variations of enhancer in ovarian cancer**. *Oncotarget* 2017, **8**(53):91558-91567.
13. Zhao M, Zhao Z: **Concordance of copy number loss and down-regulation of tumor suppressor genes: a pan-cancer study**. *BMC Genomics* 2016, **17 Suppl 7**:532.
14. Zhao M, Kim P, Mitra R, Zhao J, Zhao Z: **TSGene 2.0: an updated literature-based knowledgebase for tumor suppressor genes**. *Nucleic Acids Res* 2016, **44**(D1):D1023-1031.
15. Cavazos DA, Brenner AJ: **Hypoxia in astrocytic tumors and implications for therapy**. *Neurobiol Dis* 2016, **85**:227-233.
16. Monteiro AR, Hill R, Pilkington GJ, Madureira PA: **The Role of Hypoxia in Glioblastoma Invasion**. *Cells* 2017, **6**(4).
17. Albulescu R, Codrici E, Popescu ID, Mihai S, Necula LG, Petrescu D, Teodoru M, Tanase CP: **Cytokine patterns in brain tumour progression**. *Mediators Inflamm* 2013, **2013**:979748.
18. Gentles AJ, Newman AM, Liu CL, Bratman SV, Feng W, Kim D, Nair VS, Xu Y, Khuong A, Hoang CD *et al*: **The prognostic landscape of genes and infiltrating immune cells across human cancers**. *Nat Med* 2015, **21**(8):938-945.
19. Tan C, Liu L, Liu X, Qi L, Wang W, Zhao G, Wang L, Dai Y: **Activation of PTGS2/NF-kappaB signaling pathway enhances radiation resistance of glioma**. *Cancer Med* 2019, **8**(3):1175-1185.

20. de Leeuw R, McNair C, Schiewer MJ, Neupane NP, Brand LJ, Augello MA, Li Z, Cheng LC, Yoshida A, Courtney SM *et al*: **MAPK Reliance via Acquired CDK4/6 Inhibitor Resistance in Cancer**. *Clin Cancer Res* 2018, **24**(17):4201-4214.

Figures

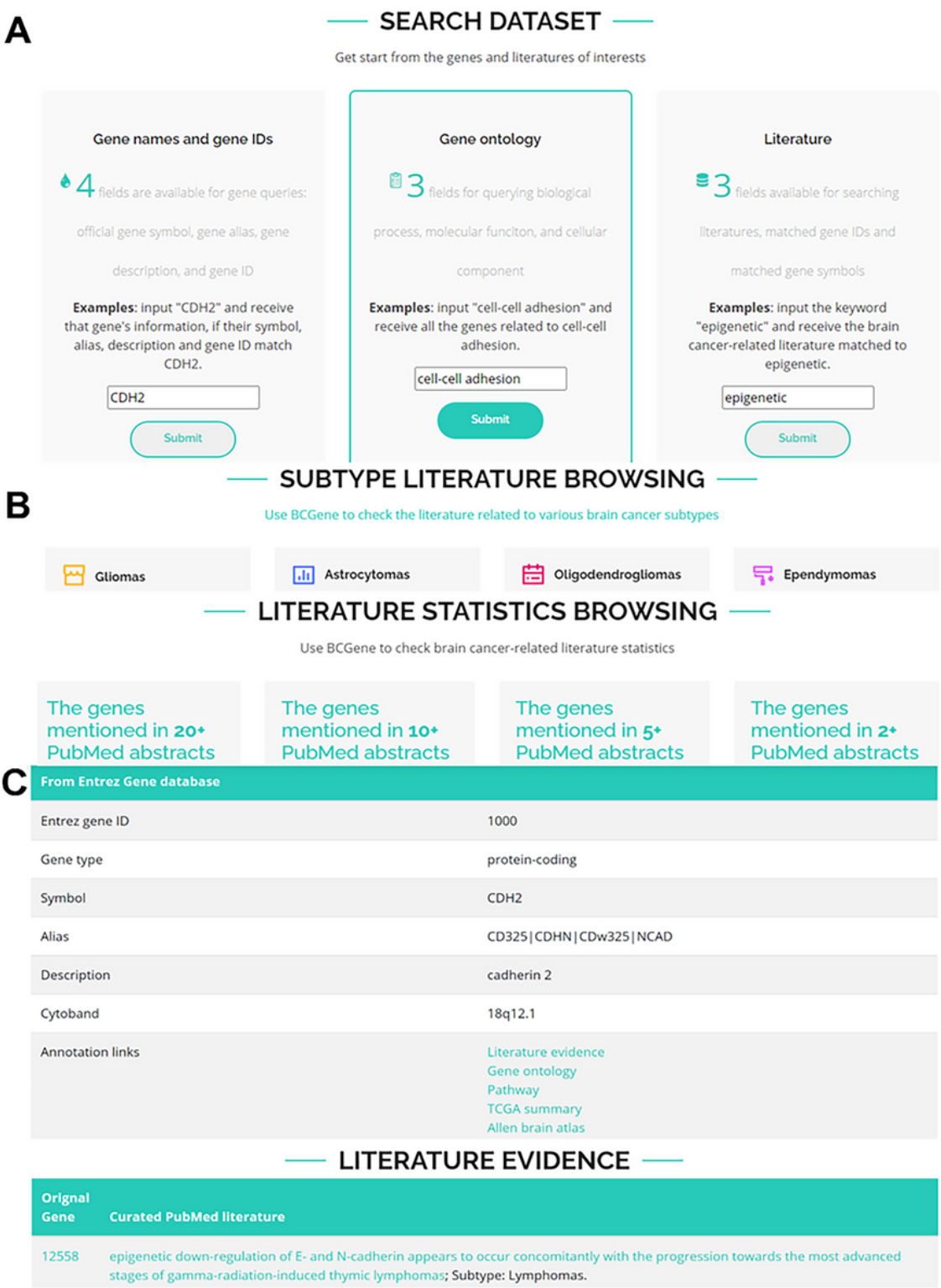

Figure 1

The BCGene database web interface. (A) Keyword-based query interface. (B) Browsing genes and literature using cancer subtypes. (C) Basic annotations and associated literature mentioning human genes in BCGene.

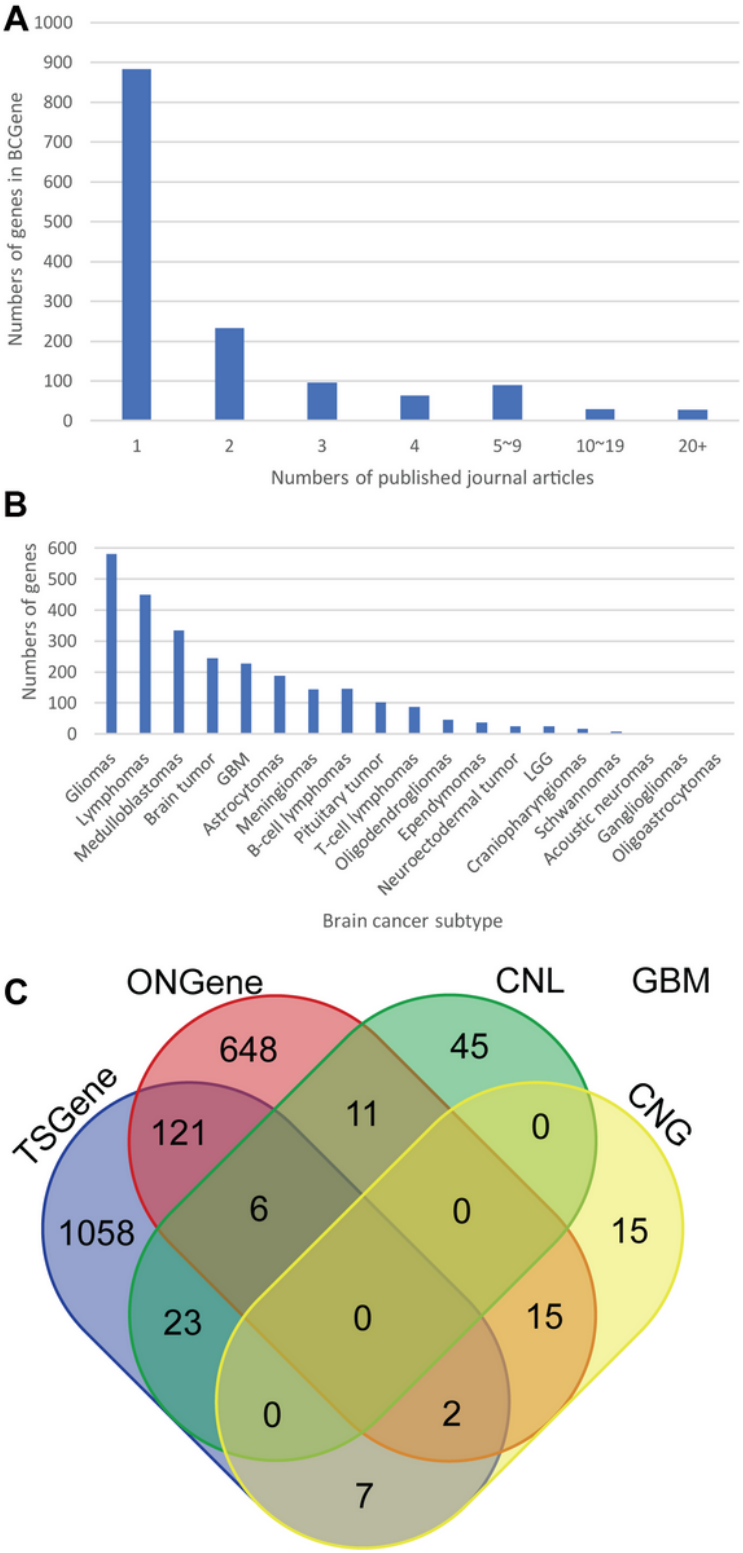

Figure 2

Overall statistics. (A) The distribution of the numbers of published articles related to all brain cancer genes in the database. (B) The numbers of genes in each subtype. (C) Venn diagram of the numbers of

potential tumor suppressors (TSGene) and oncogenes (ONGene) for glioblastoma (GBM). CNL, copy number loss; CNG, copy number gain.

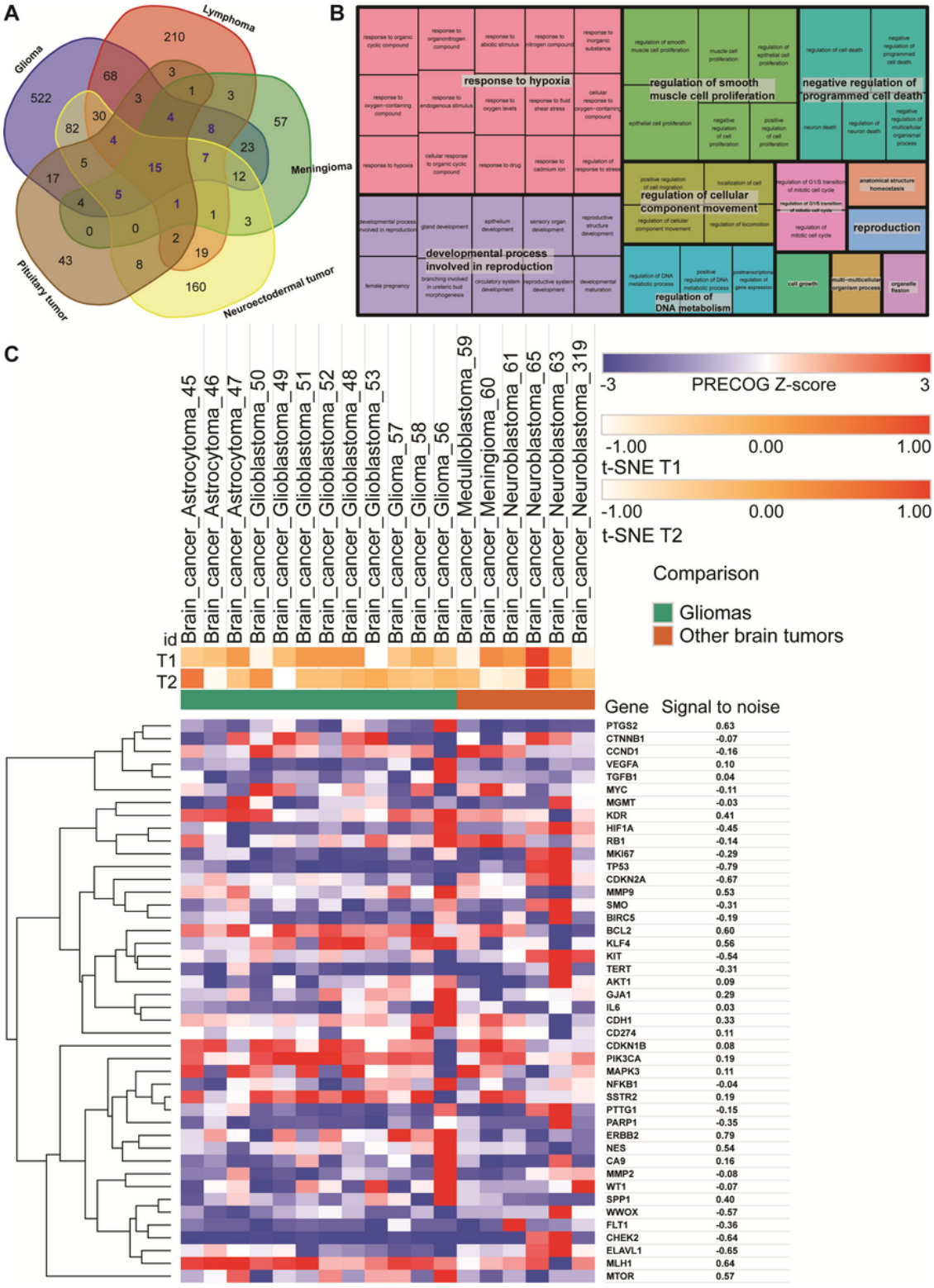

Figure 3

Overlapping and functional enrichment for genes associated with different subtypes. (A) Venn diagram of known genes from different subtypes. (B) Gene ontology enrichment analysis of the 44 genes shared by multiple subtypes. (C) Heatmap of the prognostic z-scores of 44 genes in the 18 brain cancer datasets.

Z-scores obtained from the PRECOG database are represented by the scale bar. The polarity of the prognostic z-scores reflect the direction of the association. t-distributed stochastic neighbor embedding (t-SNE) analysis characterized the similarities among datasets. Signal-to-noise ratios measured the levels of both signal and background noise. A positive ratio indicates that the signal was stronger than the noise.

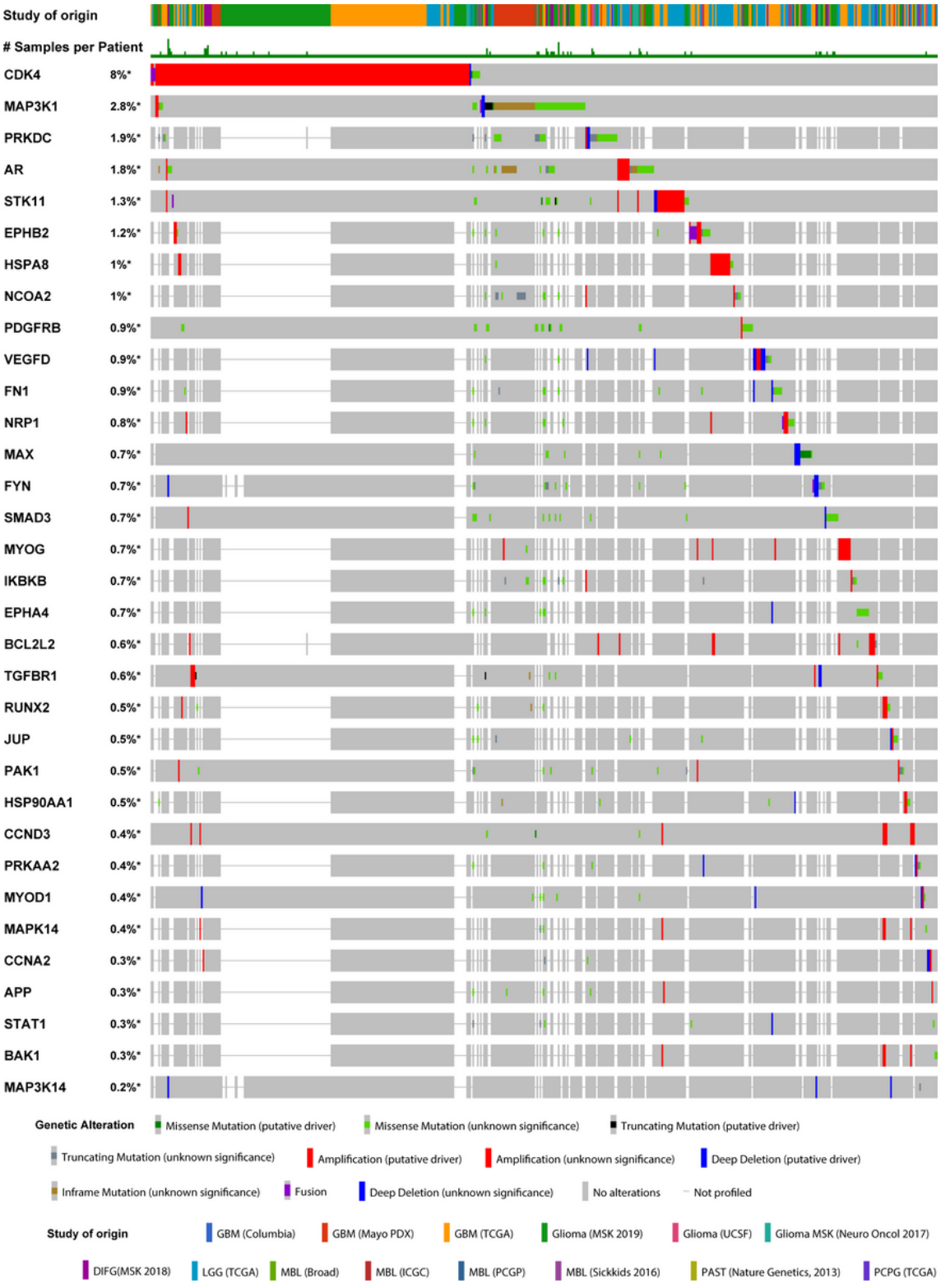

Figure 4

The top ranked genes' sample-based mutational profiles across 2,997 patient samples from 14 different brain cancer studies.

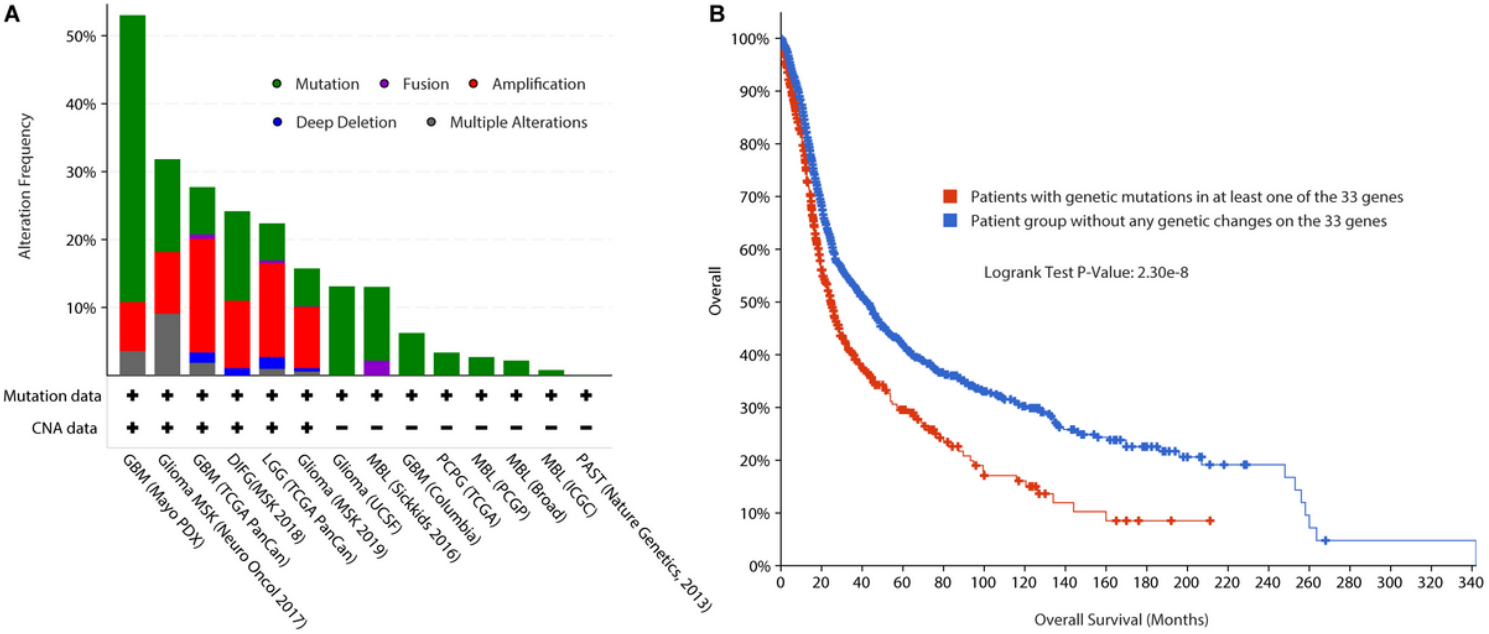

Figure 5

Mutation of the 33 top-ranked genes and overall patient survival. (A) The mutational frequency comparison of the 33 genes in 14 brain cancer studies. (B) The overall patient survival analysis of 2,997 samples from the 14 studies.

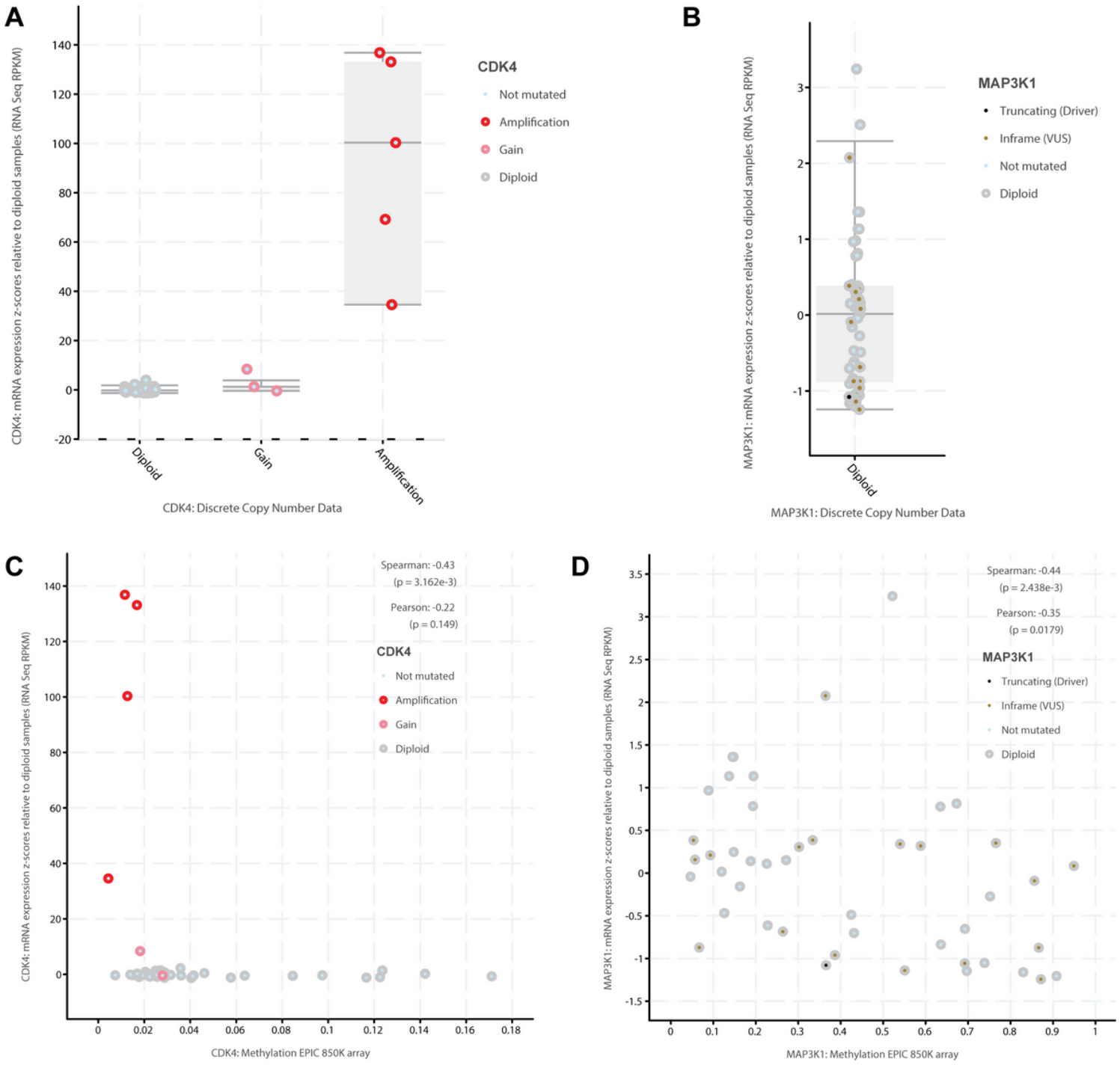

Figure 6

Correlation analyses for CKD4 and MAP3K1. (A) and (B) The relationship between copy number changes and matched mRNA expression for CDK4 (A) and MAP3K1 (B). Data are means with standard errors bounded by the gray boxes and the whiskers are the 95% confidence interval. (C) and (D) The relationships between DNA methylation and matched mRNA expressions for CDK4 (C) and for MAP3K1 (D).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryInformation.docx
- TableS1BCGene.xlsx
- TableS2subtype.xlsx
- TableS3GeneRank.xlsx